Information and Knowledge Management (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/infkm.2023.040109
ISSN 2523-5893 Vol. 4 Num. 1

Application of Big Data Analysis in Personalized Service Management of University Libraries

Mingde He*

Shandong Jiaotong Vocational College, Weifang, Shandong, 261206, China
**corresponding author*

Keywords: Big Data, Library Management, Personalized Service, Data Analysis Algorithm

Abstract: Big data (BD) is currently the most popular technology. Information technology based on BD has become a new technology that can support large-scale data processing. Its application scope is mainly in the comprehensive processing of large-scale information, and those university libraries that need large-scale data processing do need corresponding technical support. Especially in efficient libraries, because it involves tens of thousands of books of college teachers and students, in the current situation, traditional library management methods are increasingly unable to adapt to the impact of massive data. Therefore, in the management of libraries, the introduction of BD technology can improve the management efficiency of libraries, thereby providing better services to more students and students. This paper first introduced the basic framework of BD. Then, BD was applied to personalized service management in university libraries, which improved the overall efficiency. Finally, the questionnaire used to survey library staff and library customers could indicate that people had a high demand for personalized services. Therefore, making good use of BD for personalized service management in university libraries is a very worthy research topic.

1. Introduction

Under the background of China's efforts to promote "Internet plus", China has entered the era of information technology for all, and the idea of BD has gradually become popular [1]. Broadly speaking, BD refers to a new information technology based on massive amounts of information and used to mine, statistics, analyze, and use data [2]. With the increasing attention paid to BD in society, BD technology has also undergone rapid development and is increasingly changing human life [3]. In 2013, the Ministry of Culture issued a document proposing to introduce advanced data technology to create a digital cultural service network with convenient services and full media coverage, and ultimately achieve a new format of library services under the network [4]. How to use data resources to innovate library management and improve its utilization value is a problem that relevant management departments in universities should pay attention to [5].

Currently, the BD analysis market is facing unprecedented market demand and development opportunities. Aiming at the utilization of BD analysis in personalized service management of university libraries, Qi S continued to explore the service and management construction in the informatization construction of university libraries based on the Internet and the era. He strove to

provide good support for the sustainable and stable development of university libraries in the new era, and further promote the modernization of library management and service [6]. In personalized services, many consumers communicate with each other through various information and communication technology tools in social networks and power networks. Research by Sirhan A A showed that the development of personalized digital libraries using BD was based on the protection of new technologies, which could provide better services to users without worrying about protecting the privacy of publishers [7]. BD can help various industries comprehensively improve their own quality and work efficiency from various aspects. Therefore, Novikov S V proposed that the role of BD technology was to become a mobile product, which was a necessary condition for improving corporate profitability through personalized customer service and predictive analysis [8]. However, the analysis of personalized service management in university libraries in the above research is not in-depth, and the analytical methods used do not incorporate objective factors, requiring improvement.

This article first defined BD, conducted an in-depth discussion of personalized service management in university libraries, and analyzed the current situation of personalized service management in university libraries using existing materials, thereby laying a theoretical foundation for the application of BD in personalized service management in university libraries.

2. Basic Description of BD

2.1. Basic Characteristics of BD

BD is a huge data resource with large amounts of data, multiple types of data, rapid growth, and the need for new data processing methods to achieve its application value. Technically, BD cannot be processed by traditional collection, storage, and processing methods [9]. BD typically consists of five basic characteristics: data volume, variety of types, rapid data growth, high data value, and high data accuracy. Specific features are shown in Figure 1.

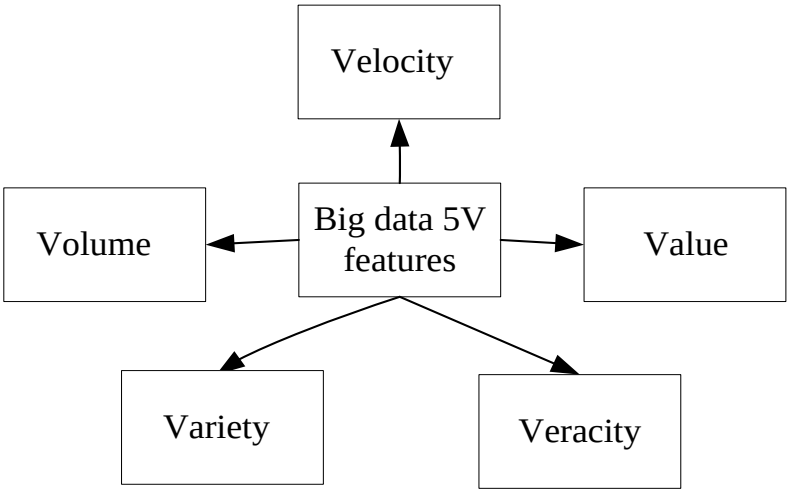

Figure 1: Basic characteristics of BD.

2.2. BD Personalization Technology

The research results of this project provide strong support for large-scale data processing. Distributed file systems represented by Hadoop are used for storage computing [10]. Compared to

traditional file systems, its storage method is different, which is stored on multiple nodes in a local area network. Moreover, there is usually only one or more central nodes to manage the indexing of file locations. When a client reads or writes data within a cluster, it first obtains the location of the file from the central node, and then communicates with the internal nodes of the cluster. The client uses the network to read or write to each node on each node. This process achieves data redundancy storage, large file segmentation, intermediate network communication, and data error recovery management through HDFS. The client can simply call according to the interface provided by HDFS, which is very convenient.

2.3. BD Analysis Algorithm

Quadratic Sorting Method

The secondary sorting method refers to sorting the values associated with a key during the reducer stage, also known as value conversion. The MapReduce framework automatically sorts the keys generated by the mapper. This indicates that the data obtained must be in order before the reducer phase is ready for execution. However, the values passed into each reducer are not sequential, and the order of the values cannot be determined (depending on the processing order of the resource provisioning process). In actual requirements, sorting values is very common. To solve this problem, it needs to use the design pattern of secondary sorting. For ease of understanding, the general MapReduce processing formula is defined as follows:

$$map(key_{1},value_{1}) => list(key_{2},value_{2}$$$reduce(key_2),list(value_{2}) => list(key_{3},value_{3})$$ (1)

Subsequently, these two formulas are explained. First, the map function accepts a k1-v1 and then outputs any number of k2-v2 pairs. Next, the reduce function receives another k-list (v) as input, and then processes it to output another k-v.

Obviously, the {v1, v2, vn...} in the input list (value_ {2}) of the reduce function is unordered, and the purpose of secondary sorting is to make them orderly. Therefore, the formula for secondary sorting can be defined according to the pattern of the formula above, as shown below:

$$map(key_{1},value_{1}) => list(key_{2},value_{2}$$$$sort(V_{1},V_{2}....V_{n}) => ({S_{1},S{2}}...S_{n}) (2)

Here, $$V represents an unordered variable, and S represents an ordered variable.

Correspondence Analysis Algorithm

Correspondence analysis algorithm is an analysis method that uses principal component analysis to describe the correlation between two or more classification variables at various levels, and its analysis results are represented by correspondence analysis diagrams that reflect the interrelationships between variables. If feature A and feature B are independent of each other, the probability of the occurrence of the ith state of feature A and the jth state of feature B simultaneously should be equal to the product of the occurrence probability of the ith state of feature A and the occurrence probability of the jth state of feature B. Therefore, there is $\overline{P_{tj}} = p_i * P_j \bullet$ ways to determine whether feature A and feature B are independent by studying the

difference between the actual probability P_{tj} and the expected probability $\overline{P_{tj}}$ of the occurrence of the ith state of feature A and the jth state of feature B.

Statistic X^2 is constructed based on the above assumptions:

$$X^2 = n\sum_{i}^{x}\sum_{j}^{m}\frac{(P_i - P_i P_j)^2}{P_i P_j} \tag{3}$$

This statistic reflects the total weighted dispersion between the observed and theoretical values of all elements. It can be proved that when n is large enough and the original assumption is H_0, X_2 follows a X_2 distribution with a degree of freedom of (n-1) (m-1).

3. University Library Personalized Service Management Experiment

This article used a questionnaire survey method to investigate the satisfaction and intention needs of library staff and library users in Area C. In the questionnaire survey, there was no communication between participants to ensure the objectivity and fairness of the questionnaire.

3.1. Questionnaire Design

A survey was conducted on the demand level for personalized service management in school libraries in the C district library, using an anonymous questionnaire. Firstly, it is necessary to determine whether the investigator is a library worker or library user in Area C, and conduct different surveys on them.

3.2. Questionnaire Distribution and Recovery

First, a designed anonymous questionnaire was printed out and distributed in paper form to library staff and library users, each with 60 copies, totaling 120 copies. A total of 105 questionnaires were collected, including 50 staff questionnaires and 55 user questionnaires. The collected questionnaires were summarized and those with unclear or unclear answers were eliminated. Finally, 94 valid questionnaires were collected, including 46 for internal staff and 48 for users.

3.3. Statistical Data Results

The collected staff questionnaires were statistically analyzed to determine the impact of the following BD on the personalized service management of university libraries. Among them, A represents a significant demand, B represents a demand, C represents a moderate demand, and D represents no demand. The data obtained during statistics are shown in Tables 1 and 2.

Table 1: Staff feedback demand level.

Degree of demand	A	B	C	D
Personalized demand for BD	32%	57%	10%	1%

Table 2: Degree of library customer feedback demand.

Degree of demand	A	B	C	D
Personalized demand for BD	55%	34%	2%	9%

From the above data, it can be seen that both library staff and library customers have a high demand for personalized services, indicating that BD can play a significant role in the management of personalized services in university libraries, improving efficiency in all aspects.

4. Results and Discussions of Personalized Service Management in University Libraries

4.1. Current Situation of Personalized Services in University Libraries

Firstly, this article studied the personalized service management of university libraries in C district. The data adopted BD in the first half of 2021 to compare the flow of library users who used personalized services in university libraries with the flow of university libraries who did not use BD. The specific data is shown in Figure 2.

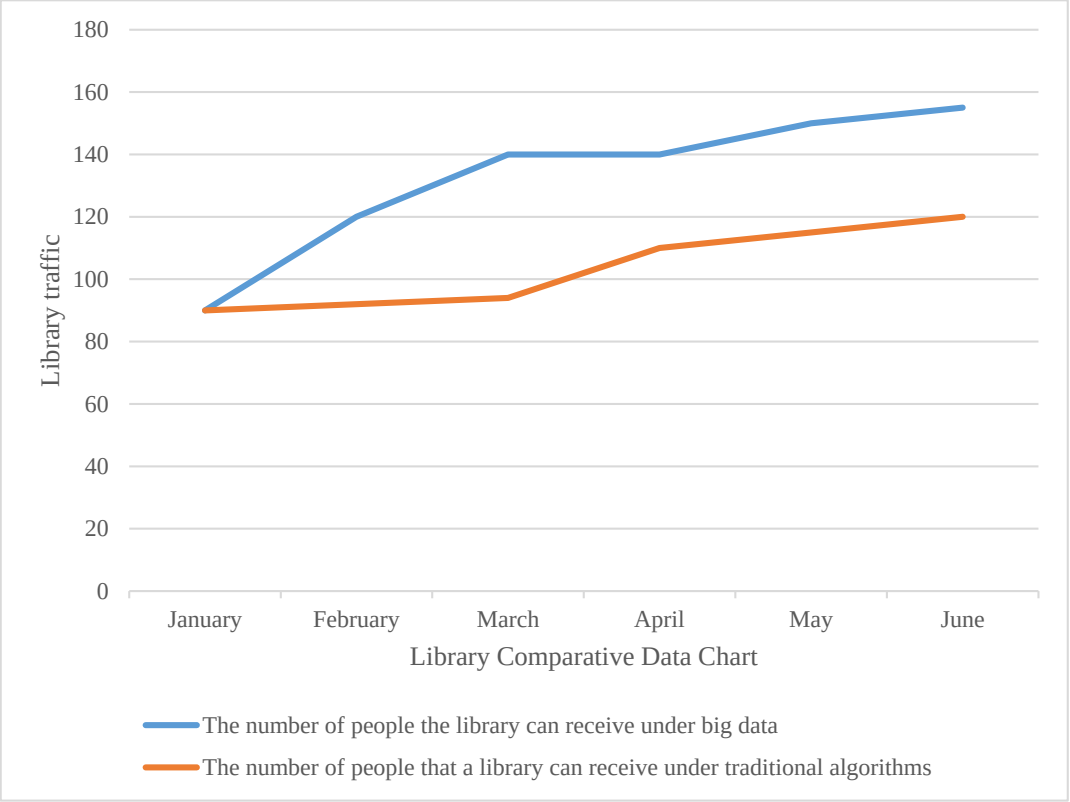

Figure 2: Comparative chart of university library passenger flow in the first half of 2021.

The data indicates that the flow of people in libraries that the use of BD to provide personalized services to university libraries is greater than the flow of people in libraries that do not use BD to provide personalized services. This indicates that BD can play a very helpful role in personalized services, and also indicates that people have a high demand for personalized management.

4.2. Self Defects of BD in Personalized Service Management of University Libraries

Currently, both in terms of data architecture and data processing, there are enormous challenges. According to industry information network research, China's existing BD analysis architecture has problems such as poor scalability, inefficient resource utilization, cumbersome application deployment, high operating costs, and high energy consumption, as shown in Figure 3 for details.

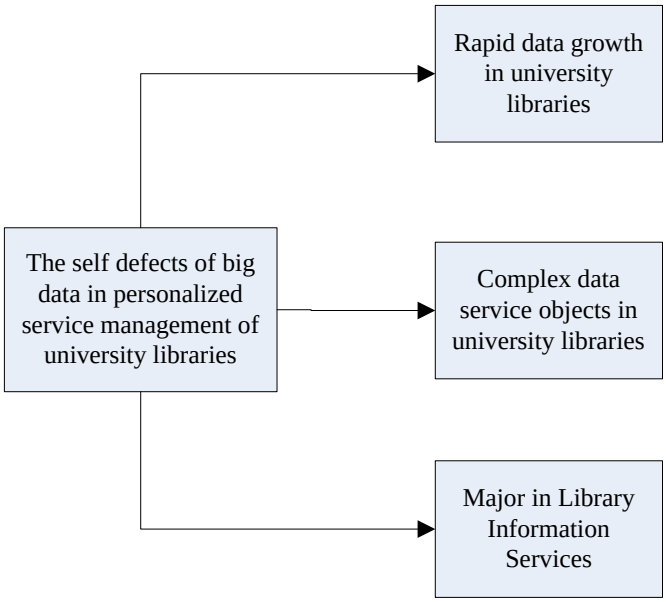

Figure 3: The inherent defects of BD in personalized service management of university libraries.

4.2.1. Rapid Data Growth in University Libraries

In the context of "BD", with the development and upgrading of various industries, data and information in various industries also take on increasingly complex forms and structures [11-12]. At the same time, various structured and semi structured data, different search methods, databases and information storage, as well as some unstructured data, can change the service mode of university libraries, improve the quality of information services in university libraries, making it an important asset of university libraries. In the era of BD, the ability of university libraries to store data information needs to be gradually improved. University libraries need to face the complex tasks of computing and processing a large number of data information within themselves. When faced with such problems, BD algorithms are the best answer.

4.2.2. Complex Data Service Objects of University Libraries

With the development of BD technology, the past information society transforms into the present knowledge society. Emerging libraries such as virtual libraries eventually become the development direction of future libraries. The data service objects of university libraries become increasingly complex. It can fully reflect the service concept of "anytime, anywhere" [13].

4.2.3. Library Information Service

Currently, BD technology has been increasingly used in social life, which can effectively ensure the correctness of data and further improve the analysis and processing capabilities of massive data [14]. For example, before the advent of cloud computing, traditional computers are unable to

effectively process massive, unstructured, complex, highly complex, and unstructured data [15]. However, with the growth of cloud computing technology, unconventional and unstructured data can be efficiently stored and processed without being limited to time and space. To this end, BD and cloud computing are closely connected, and the organic integration of the two is expected to improve the efficiency of BD collection.

4.3. Suggestions for Correcting Defects in Personalized Service of University Libraries

4.3.1. Strengthening the Combination and Integration of Information Resources

In the BD environment, the types of data information contained in university libraries have also shown a diversified development trend, including a large amount of structured data information, semi structured data information, and unstructured data information. If libraries can adopt certain methods for such chaotic data information and classify it according to relevant rules, it is more convenient for users to use [16]. During the process of accessing the library website and interacting with the library, users generate a large number of data information. Using BD mining and analysis technology can better organize and integrate these massive data information [17].

4.3.2. Deeply Mining Information Resource Content

To provide targeted services to readers, it is necessary to first have the right to know about the readers, and to learn the demands of the readers [18]. User information actions include user information requests and user information actions. University libraries can obtain basic information about readers based on their registration information, interaction information with readers, etc. [19]. Through data mining and data analysis, users' needs for information can be obtained. Using BD technology, university libraries can comprehensively and real-time record user information behavior, assist universities in building user information demand models, and ensure the personalized service quality of university libraries [20].

5. Conclusions

To break through the current dilemma of BD, there are three key points. Firstly, the system construction and data resource sharing system and mechanism have been established and improved. The second is to provide relevant technical support, data resource inheritance and management tools, data analysis software, and visual information. Data needs to be absorbed, planned, and analyzed to find ways to give greater value, with a focus on personalization. At the same time, university libraries actively respond to and participate in database construction and development with their own advantages. Implementing personalized services has become an indispensable part of today's university libraries in responding to challenges, marking a significant change in service concepts and operational models. BD technology can excavate potential information and knowledge from a large amount of data, providing technical support for the personalized service of university libraries. In today's China, BD has been widely utilized in many aspects, such as business intelligence, government decision-making, public services, and so on.

References

[1] Pisareva O M. Analysis of the State and Characteristics of the Development Potential of Strategic Planning Tools in the Digital Transformation Conditions of the Economy and Management. MIR (Modernization Innovation Research), 2019, 9(4):502-529.

[2] Zhang F, Liu Q H. Innovation of Geography by Artificial Intelligence _Big Data_Cloud Computing. Journal of Global Change Data & Discovery, 2018, 2(3):362-365.

[3] Clim A, Zota R D, Tinica G. Big Data in home healthcare: A new frontier in personalized medicine. Medical emergency services and prediction of hypertension risks. International Journal of Healthcare Management, 2019, 12(3):241-249.
[4] Manikyam N, Kumar S M. Methods and techniques to deal with big data analytics and challenges in cloud computing environment. International Journal of Civil Engineering and Technology, 2017, 8(4):669-678.
[5] Lv X, Li M. Application and Research of the Intelligent Management System Based on Internet of Things Technology in the Era of Big Data. Mobile Information Systems, 2021, 2021(16):1-6.
[6] Qi S. Approaches to Information Service and Management Construction in University Libraries. International Journal of Social Science and Education Research, 2019, 2(5):41-45
[7] Sirhan A A, Abdrabbo K M, Tawalbeh S, et al. Digital rights management (DRM) in libraries of public universities in Jordan. Library management, 2019, 40(8/9):496-502.
[8] Novikov S V. Data Science and Big Data Technologies Role in the Digital Economy. TEM Journal, 2020, 9(2):756-762.
[9] Bai X, Li J. Personalized dynamic evaluation technology of online education quality management based on artificial intelligence big data. Journal of Intelligent and Fuzzy Systems, 2021(3):1-10.
[10] Savitskaya T E. New Library Services in the Framework of Digital Humanities Projects: Foreign Experience. Bibliotekovedenie [Russian Journal of Library Science], 2021, 70(1):55-64.
[11] Laterza V. Could Cambridge Analytica Have Delivered Donald Trump's 2016 Presidential Victory? An Anthropologist's Look at Big Data and Political Campaigning. Public Anthropologist, 2021, 3(1):119-147.
[12] Qin J, Zhou Z, Tan Y. A Big Data Text Coverless Information Hiding Based on Topic Distribution and TF-IDF. International Journal of Digital Crime and Forensics, 2021, 13(4):40-56.
[13] Kazakova N, Mel'Nik M, Dudorova E. Prospects for Implementing Big Data Analytics into the Auditing Profession. Auditor, 2021, 7(3):40-47.
[14] Xie C, Gui Y, Wong C S. Converting Double-Stranded DNA-Encoded Libraries (DELs) to Single-Stranded Libraries for More Versatile Selections. ACS Omega, 2022, 7(13):11491-11500.
[15] Hohnen P, Ulfstjerne M A, Krabbe M S. Assessing Creditworthiness in the Age of Big Data: A Comparative Study of Credit Score Systems in Denmark and the US. Journal of Extreme Anthropology, 2021, 5(1):29-55.
[16] Deng Y. Construction of higher education knowledge map in university libraries based on MOOC. The Electronic Library, 2019, 37(5):811-829.
[17] Ye P, Liu L, Gao L. Influence of Information and Service Quality on Users' Continuous Use of Mobile Libraries in China. Journal of Cases on Information Technology, 2020, 22(1):57-71.
[18] Savitskaya T E. New Library Services in the Framework of Digital Humanities Projects: Foreign Experience. Bibliotekovedenie [Russian Journal of Library Science], 2021, 70(1):55-64.
[19] Prakash P, Singh M P. Disaster Management Strategies of the Central University Libraries in Uttar Pradesh, India: A Study. GeroFam-A peer reviewed evidence-based gerontology-oriented family practice journal, 2021, 10(40):147-152.
[20] WenjinZHANG, ZhongyuQIN, ZhaoFENG. Big data analysis for detection of web brute-force attack. Journal of Shenzhen University Science and Engineering, 2020, 37(Z1):44-49.